

Efficient and robust estimation for financial returns: an approach based on q -entropy

Davide Ferrari^{a,*}, Sandra Paterlini^a

^a *Dept. of Economics, CEFIN and RECent, University of Modena and Reggio Emilia, Italy*

Abstract

We consider a new robust parametric estimation procedure, which minimizes an empirical version of the Havrda-Charvát-Tsallis entropy. The resulting estimator adapts according to the discrepancy between the data and the assumed model by tuning a single constant q , which controls the trade-off between robustness and efficiency. The method is applied to expected return and volatility estimation of financial asset returns under multivariate normality. Theoretical properties, ease of implementability and empirical results on simulated and financial data make it a valid alternative to classic robust estimators and semi-parametric minimum divergence methods based on kernel smoothing.

Keywords: q -entropy, robust estimation, power-divergence, financial returns

JEL: C13, G11

*Department of Economics, University of Modena and Reggio Emilia, viale Berengario 51, Modena, 41100, Italy. Phone: +39 059 205 6838, Fax: +39 059 205 6947, *E-mail address:* davide.ferrari@unimore.it

1. Introduction

Many decision-making and asset pricing models in finance rely on assumptions on the stochastic model underlying the data. The multivariate normal assumption is one of the most popular, as mean and covariance estimates are sufficient for computing the Markowitz's mean-variance (MV) optimal portfolio allocation (Markowitz, 1952). However, for a typical sample of financial returns, the empirical distribution deviates in various amounts from normality. Figure 1 shows a normal quantile plot for the Standard & Poor's 500 (S&P500) log-return. While the bulk of the observations follows normality fairly closely, many values in the tails are far to meet such an assumption. Between these two extremes, there is an interesting portion of observations, which departs from normality in various degrees.

Statistical regularities of financial returns, such as the leptokurtic non-robust behavior, volatility clustering and the asymmetry gain/losses (Cont, 2001), have often lead to discard the normal model in favor of more sophisticated representations (Bauwens et al., 2006; Zhao, 2008). However, an increased model complexity adds huge costs in terms of interpretability, stability of parameter estimates and model calibration (Zhao, 2008). A valid alternative is to retain a simple stochastic model, while making the estimation method sensitive to the amount of information carried by each observation relative to the assumed model. Under this viewpoint, the classic maximum likelihood estimator (MLE) equally weights each observation, it is highly non-robust and some of its desiderate properties – such as efficiency – are not trustworthy in presence of deviations of the data from the assumed model (Huber, 1981; Hampel et al., 1986). In practice, bad estimates translate into

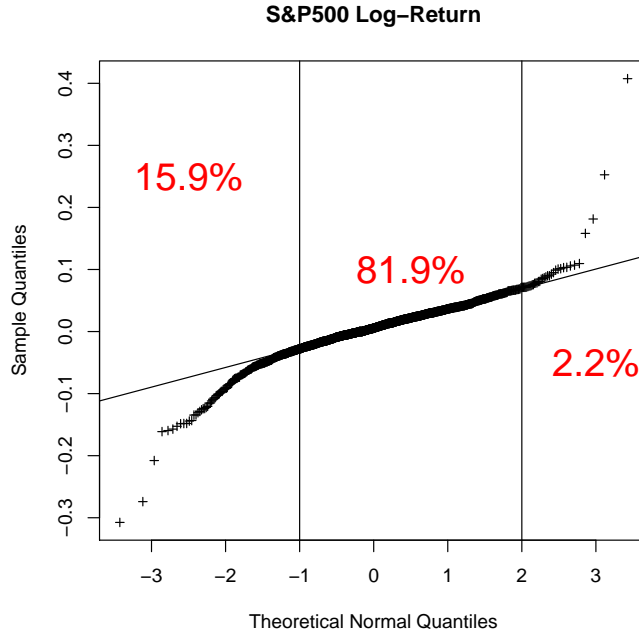


Figure 1: QQ-plot of monthly log-returns of the S&P500 stock index (Period: January 1871- August 2008).

undesirable decisions of asset allocation due to the asset weights sensitivity to the estimates (Best and Grauer, 1991). Instead, classic robust methods and minimum-divergence based methods allow for discriminating the amount of information carried by each observation, thus providing reasonable estimates even in presence of deviations from the assumptions.

Much work has been devoted to construct portfolios not overly affected by deviations of the data from the multivariate normal model. One line of research aimed to improve the stability of portfolio weights by employing classic robust estimators of the mean and variance. These contributions differ mainly for the type of estimator employed: Vaz-de Melo and Camara (2005)

use M-estimators, Perret-Gentil and Victoria-Feser (2005) use S-estimators and Welsch and Zhou (2007) use minimum covariance determinant (MCD) and winsorization. Lauprete et al. (2002) perform parameter estimation and portfolio optimization in a single step based on M-estimation of the covariance matrix. DeMiguel and Nogales (2009) consider portfolios based on both M- and S-estimators and provide analytical bounds for the sensitivity of the investment strategy to changes in the parametric assumptions. Although these contributions have the merit to address the role of robust estimation for improving MV portfolios, traditional robust estimators still suffer dramatic losses of efficiency compared to maximum likelihood. This issue is crucial in multivariate problems with a large number of parameters.

In a different direction, our work is developed within a minimum divergence framework, i.e. considering minimization of some appropriate data-based divergence between an assumed model and the true model density underlying the data. Depending on the choice of the divergence, minimum divergence estimators can afford considerable robustness at minimal expense of efficiency. Beran (1977) was a pioneer of divergence methods for robustness, putting forward the well-known Minimum Hellinger Distance Estimator (MHDE), which can tolerate about 50% of bad data and yet maintaining first-order efficiency. Subsequent developments include those of Basu and Lindsay (1994) and Lindsay (1994). All the above approaches, however, require non-parametric density estimation, which is troublesome in multi-dimensional problems.

In this paper, we consider an estimator of location and scale obtained by minimizing a family of quasi-logarithmic density divergences. The methodol-

ogy is motivated from an information-theoretical perspective, since it amounts to minimize an empirical version of Tsallis-Havrda-Charvát (HCT) entropy (Havrda and Charvát, 1967; Tsallis, 1988) or q -entropy. The resulting estimator, named Maximum Lq -Likelihood Estimator (ML q E) was introduced by Ferrari and Yang (Ferrari and Yang, 2009) in the context of small probability estimation. A Fisher-consistent version of ML q E and its infinitesimal robustness properties are examined by Ferrari and La Vecchia (2009).

The ML q E is appealing as it conciliates efficiency and robustness aspects, usually requiring distinct techniques. The former is prioritized when the model is thought to be appropriate for the data at hand, while the latter is stressed when it is not. The behavior of the ML q E depends on a single parameter q , which controls such a trade-off. When the data are consistent with normality – or other model specification – and $q \rightarrow 1$, the ML q E corresponds to the MLE. When $q < 1$, the estimator gains robustness, yet maintaining considerable efficiency. If $q = 1/2$, the ML q E is a minimizer of a version of the Hellinger distance, which has the perk of not involving degrees of nonparametric analysis. This aspect is valuable as it avoids the difficulties related to bandwidth selection in multiple dimensions, which is instead required by MHDE.

The estimator can be applied to any parametric family. An important feature of the ML q E is that the extent to which each observation is an outlier is determined in terms of the model itself. The method relies on minimizing a weighted version of the log-likelihood function, where the weights are proportional to the $(1 - q)$ th power of the assumed density. As a consequence, a simple and fast algorithm based on a re-weighting strategy for computing

ML q estimates is provided. For the multivariate normal case, the steps of the algorithm reduce to a simple variable transformation. A fast procedure for the optimal choice of q is proposed, making the new method attractive not only for its theoretical properties but also for its ease of implementability and fast convergence.

Section 2 describes the q -entropy minimization approach and the location/scale estimator. In Section 3, we provide the asymptotics and robustness properties of the estimator. In Section 4, we report empirical results on simulated data. In Section 5, we focus on financial data and MV portfolio allocations. In Section 6, we discuss our findings and suggest future research directions.

2. q -entropy minimization

Let \mathcal{G} be the class of all distributions with pdf g and support $\mathcal{X} \in \mathbb{R}^p$, $p \geq 1$ with respect to the Lebesgue measure. Further, let $\mathcal{F}(\Theta)$, $\Theta \in \mathbb{R}^k$, $k \geq 1$ be a parametric family of target distributions with densities f on \mathcal{X} . The distribution generating the data is regarded as close but not exactly equal to some member of $\mathcal{F}(\Theta)$. Here, we consider f to be a multivariate normal density with mean vector $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$. The current approach, however, can be applied to other target models. For example, t -Student and stable Paretian distributions are common alternatives in finance to capture the leptokurtic behavior of the financial returns. The overall parameter vector as $\boldsymbol{\theta}' = (\boldsymbol{\mu}', \text{vech}(\boldsymbol{\Sigma})') \in \mathbb{R}^k$, where $k = p + p(p + 1)/2$.

Consider the power divergence between $g(\mathbf{x})$ and $f(\mathbf{x}; \boldsymbol{\theta})$:

$$D_q(\boldsymbol{\theta}; g) = -\frac{1}{q} \int g(\mathbf{x}) L_q \left\{ \frac{f(\mathbf{x}; \boldsymbol{\theta})}{g(\mathbf{x})} \right\} d\mathbf{x}, \quad (1)$$

where $L_q(u) = (u^{1-q} - 1)/(1 - q)$, $u > 0$, if $q \in (-\infty, \infty) \setminus \{1\}$. When $q = 1$, the integrand is undefined and we set $\log(\cdot) = \lim_{q \rightarrow 1} L_q(\cdot)$, recovering the Kullback-Leibler (KL) divergence

$$D_1(\boldsymbol{\mu}, \boldsymbol{\Sigma}; g) = \int \frac{1}{2} (\mathbf{x} - \boldsymbol{\mu})' \boldsymbol{\Sigma}^{-1} (\mathbf{x} - \boldsymbol{\mu}) g(\mathbf{x}) d\mathbf{x} - \frac{1}{2} \log |\boldsymbol{\Sigma}| + \text{const.}$$

By the Law of Large Numbers, given data \mathbf{x}_i , $i = 1, \dots, n$, the above expectation is approximated by the empirical distribution of the data. The minimizer is the MLE, which is optimal when $g = f$. For financial returns, however, since such a requirement is not met, asymptotic consistency and efficiency of the MLE are unreliable.

If $q = 1/2$, (1) is twice the Hellinger distance: $2 \int [f^{1/2}(\mathbf{x}; \boldsymbol{\theta}) - g^{1/2}(\mathbf{x})]^2 d\mathbf{x}$. Beran (1977) considered minimizing such a quantity, introducing the semi-parametric MHDE. Later, Basu and Lindsay (1994) and Lindsay (1994) extended Beran's approach to the general case $q \neq 1$. Although these methods were shown to be highly robust and fully efficient at the model, their implementation requires non-parametric smooth estimate \hat{g}_h of g , provided a proper choice of the bandwidth h . In multivariate settings, as for correlated financial data, choosing h is often challenging. In addition, the curse of dimensionality makes this approach impracticable even for a moderate number of variables.

These issues can be avoided by approaching minimization of (1) indirectly. We consider minimization of the q -entropy functional

$$H_q(\boldsymbol{\theta}; g) = - \int g(\mathbf{x}) L_q \{ f(\mathbf{x}; \boldsymbol{\theta}) \} d\mathbf{x}. \quad (2)$$

This is equivalent to minimize $D(\boldsymbol{\theta}; g^*)$, where $g^*(\mathbf{x}) = g^{1/q}(\mathbf{x}) / \int g^{1/q}(\mathbf{x}) d\mathbf{x}$ is a power-transformed version of the true density (see Lemma 1, Ferrari and La Vecchia (2009)). Therefore, a transformation on the estimates in order to recover consistent estimates for the right target g is required. An important advantage is that Eq. (2) can be simply approximated by the Law of Large Numbers, without any density smoothing.

The previous considerations motivate the following estimating functional:

$$T(g) = \boldsymbol{\Psi}_q \left(\arg \min_{\boldsymbol{\theta} \in \Theta} \{H_q(\boldsymbol{\theta}; g)\} \right), \quad (3)$$

where $\boldsymbol{\Psi}_q \in \mathbb{R}^{k \times k}$ is a diagonal matrix such that $\text{diag}(\boldsymbol{\Psi}_q)_j = 1$ for $j = 1, \dots, p$ and $\text{diag}(\boldsymbol{\Psi}_q)_j = q^{-1}$ for $j = p+1, \dots, k$, where $k = p+p(p+1)/2$. Let $\boldsymbol{\theta}_0$ denote the unique minimizer of (1), representing the parametric density closest to g in the sense of the power divergence D_q . The transformation $\boldsymbol{\Psi}_q$ ensures Fisher-consistency, of $T(\cdot)$, i.e. $T(f(\cdot; \boldsymbol{\theta}_0)) = \boldsymbol{\theta}_0$ (Ferrari and La Vecchia, 2009).

3. A fully parametric estimator

Ferrari and Yang (2009) introduce the MLqE in the context of small tail probability estimation. In this paper, we consider the following re-centered version of the MLqE of $\boldsymbol{\theta}_0 = (\boldsymbol{\mu}'_0, \text{vech}\boldsymbol{\Sigma}'_0)$ in order to obtain Fisher-consistency.

$$\hat{\boldsymbol{\theta}}_{q,n} = \boldsymbol{\Psi}_q \left(\arg \max_{\boldsymbol{\theta} \in \Theta} \sum_{i=1}^n L_q \{f(\mathbf{x}_i; \boldsymbol{\theta})\} \right). \quad (4)$$

The above estimator entails solving the estimating equations

$$0 = \sum_{i=1}^n f(\mathbf{x}_i; \boldsymbol{\theta})^{1-q} u(\mathbf{x}_i; \boldsymbol{\theta}), \quad (5)$$

where $u(\mathbf{x}, \boldsymbol{\theta}) = \nabla_{\boldsymbol{\theta}} \log f(\mathbf{x}; \boldsymbol{\theta})$ denotes the maximum likelihood score vector. For estimating $\boldsymbol{\mu}$ the rescaling matrix $\boldsymbol{\Psi}_q$ is the identity matrix, while for $\boldsymbol{\Sigma}$ the final solution involves dividing by q . When $q = 1$, $L_q(\cdot) \rightarrow \log(\cdot)$, $\boldsymbol{\Psi}_q = \mathbf{I}_k$ and $\widehat{\boldsymbol{\theta}}_{1,n}$ is actually the MLE of $\boldsymbol{\theta}_0$. The estimator in (4) is related to the robustification strategy proposed by Windham (1995). However, Windham applies model-based re-weighting to a general estimating function. Here, we focus on the particular case where the estimating function is actually the score function.

3.1. Properties

3.1.1. Asymptotics and standard errors

For fixed q , the solution of Eq. (5) is an M-estimator and the asymptotic distribution of MLqE can be derived from existing theory (Hampel et al., 1986). Let $\mathbf{x}_1, \dots, \mathbf{x}_n$ be independent observations from g . As $n \rightarrow \infty$: (i) There exists a sequence $\widehat{\boldsymbol{\theta}}_{q,n}$ such that $\widehat{\boldsymbol{\theta}}_{q,n}$ converges to $\boldsymbol{\theta}_0$ in probability. (ii) For any consistent sequence $\widehat{\boldsymbol{\theta}}_{q,n}$, $\sqrt{n}(\widehat{\boldsymbol{\theta}}_{q,n} - \boldsymbol{\theta}_0)$ converges in distribution to a multivariate normal with zero mean vector and covariance

$$\mathbf{V}_q(\boldsymbol{\theta}, g) = \mathbf{J}_q^{-1}(\boldsymbol{\theta}, g) \mathbf{K}_q(\boldsymbol{\theta}, g) \mathbf{J}_q^{-1}(\boldsymbol{\theta}, g), \quad (6)$$

where $\mathbf{J}_q, \mathbf{K}_q \in \mathbb{R}^{k \times k}$ are defined as

$$\mathbf{K}_q(\boldsymbol{\theta}, g) = E_g[f(\mathbf{x}; \boldsymbol{\theta})^{2(1-q)} u(\mathbf{x}, \boldsymbol{\theta}) u(\mathbf{x}, \boldsymbol{\theta})'], \quad (7)$$

$$\mathbf{J}_q(\boldsymbol{\theta}, g) = -\boldsymbol{\Psi}_q E_g f(\mathbf{x}; \boldsymbol{\theta})^{1-q} [(1-q) u(\mathbf{x}, \boldsymbol{\theta}) u(\mathbf{x}, \boldsymbol{\theta})' + \nabla_{\boldsymbol{\theta}} u(\mathbf{x}, \boldsymbol{\theta})]. \quad (8)$$

The above expressions follow directly from Theorems 5.14 and 5.41 in Van der Vaart (1998). If $q = 1$ and the model is correctly specified, i.e. $g(\mathbf{x}) = f(\mathbf{x}, \boldsymbol{\theta}_0)$, one can see that $\mathbf{K}_q(\boldsymbol{\theta}_0) = \mathbf{J}_q(\boldsymbol{\theta}_0)^{-1} \mathbf{J}_q(\boldsymbol{\theta}_0)^{-1}$ and $V_q(\boldsymbol{\theta}_0)$ is just

the inverse of Fisher information matrix. Consistent estimates of \mathbf{V}_q can be obtained by computing Huber's sandwich estimator (Huber, 1981) by replacing the distribution of the data instead of g in the expressions (7) and (8) and computing $\widehat{\mathbf{V}}_q(\widehat{\boldsymbol{\theta}}_{q,n}) = \widehat{\mathbf{J}}_q(\widehat{\boldsymbol{\theta}}_{q,n})^{-1} \widehat{\mathbf{K}}_q(\widehat{\boldsymbol{\theta}}_{q,n}) \widehat{\mathbf{J}}_q(\widehat{\boldsymbol{\theta}}_{q,n})^{-1} / (n - 1)$. First order and second order derivatives characterizing the integrands can be obtained by numerical differentiation. Estimates of the variance of the MLqE and confidence intervals for smaller sample sizes can be also computed using standard re-sampling techniques such as bootstrap.

3.1.2. Local robustness

We focus on small deviations from normality by assuming that $g = g_\epsilon$ belongs to the contaminate family $(1 - \epsilon)N_p(\boldsymbol{\mu}_0, \boldsymbol{\Sigma}_0) + \epsilon\Delta_{\mathbf{x}}$, where $0 \leq \epsilon < 1/2$ represents the proportion of data from the unknown contaminating distribution and $\Delta_{\mathbf{x}}$ is Dirac's Delta function, placing the entire mass on a single point.

A useful tool to study the sensitivity of the estimator to data contamination is represented by the *influence function*, IF: $\mathbb{R}^k \mapsto \mathbb{R}$ defined as $\text{IF}(\mathbf{x}, T(g_\epsilon)) = \partial T(g_\epsilon) / \partial \epsilon |_{\epsilon=0}$ whenever the limit exists. The first-order Von Mises expansion gives an approximation of the bias as: $\text{Bias} = T(g_\epsilon) - \boldsymbol{\theta}_0 \approx \epsilon \text{IF}(\mathbf{x}, T(g_\epsilon))$. Therefore, a bounded IF implies that the estimator has also bounded asymptotic bias. A standard calculation (e.g., see Hampel et al. (1986)), gives:

$$\text{IF}(\mathbf{x}, \widehat{\boldsymbol{\theta}}_{q,n}) = -\mathbf{J}_q^{-1}(\boldsymbol{\theta}_0) [f(\mathbf{x}; \boldsymbol{\theta}_0)^{1-q} u(\mathbf{x}, \boldsymbol{\theta}_0)]. \quad (9)$$

When $q = 1$, the IF is just proportional to the score function $u(\mathbf{x}, \boldsymbol{\theta})$. In the

case of the mean and covariance estimates of $N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$,

$$\text{IF}(\mathbf{x}, \widehat{\boldsymbol{\mu}}_{1,n}) = \mathbf{x} - \boldsymbol{\mu}, \text{ and } \text{IF}(\mathbf{x}, \widehat{\boldsymbol{\Sigma}}_{1,n}) = (\mathbf{x} - \boldsymbol{\mu})(\mathbf{x} - \boldsymbol{\mu})' - \boldsymbol{\Sigma}. \quad (10)$$

Clearly, the IFs above expressions are unbounded in \mathbf{x} . Therefore, under contamination, we expect large biases. For $q < 1$, however, Eq.(9) gives bounded IFs:

$$\text{IF}(\mathbf{x}, \widehat{\boldsymbol{\mu}}_{q,n}) \propto \mathbf{d}(\mathbf{x}) \exp \left\{ -(1-q) \|\mathbf{d}(\mathbf{x})\|^2 / 2 \right\} \quad (11)$$

$$\text{IF}(\mathbf{x}, \widehat{\boldsymbol{\Sigma}}_{q,n}) \propto [\mathbf{d}(\mathbf{x})\mathbf{d}(\mathbf{x})' - \boldsymbol{\Sigma}] \exp \left\{ -(1-q) \|\mathbf{d}(\mathbf{x})\|^2 / 2 \right\}, \quad (12)$$

where $\mathbf{d}(\mathbf{x}) = \boldsymbol{\Sigma}^{-1/2}(\mathbf{x} - \boldsymbol{\mu})$. Note that for the multivariate normal, (11) and (12) define re-descending estimators, meaning that the IFs approach to zero as $\|\mathbf{x}\| \rightarrow \infty$. This is clearly seen in Figure (3.1.2), where we represent the IFs (11) and (12) up to a rescaling constant for a bivariate normal with zero mean and identity covariance matrix.

3.1.3. Global robustness: breakdown at the edge

Global robustness is assessed by computing gross-error breakdown point as defined in (Hampel et al., 1986), which is the value $0 \leq \epsilon < 1/2$ of the contamination at which the estimator still gives some relevant information about the model parameters. Consider maximizing the estimating function over the target parameters \mathbf{m} and \mathbf{S} :

$$\begin{aligned} \psi(\mathbf{m}, \mathbf{S}) &= \int g_\epsilon(\mathbf{z}) L_q \{ f(\mathbf{z}; \mathbf{m}, q^{-1}\mathbf{S}) \} d\mathbf{z} \\ &= \frac{1-\epsilon}{1-q} \int f(\mathbf{z}; \boldsymbol{\mu}, \boldsymbol{\Sigma}) f(\mathbf{z}; \mathbf{m}, q^{-1}\mathbf{S})^{1-q} d\mathbf{z} \\ &\quad + \frac{\epsilon}{1-q} \int \Delta_{\mathbf{x}}(\mathbf{z}) f(\mathbf{z}; \mathbf{m}, q^{-1}\mathbf{S})^{1-q} d\mathbf{z} - \frac{2}{1-q}. \end{aligned} \quad (13)$$

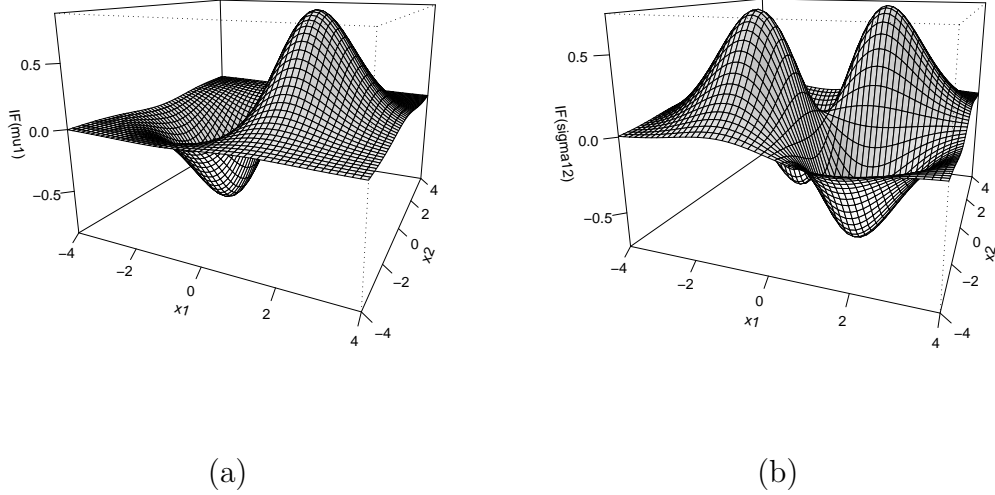


Figure 2: IFs for normal distribution $N_2(\mathbf{0}, \mathbf{I})$. (a) IF for the first component of the mean. (b) IF for an element of the covariance matrix.

To evaluate the first term in (13), we use the following result which is derived using straightforward integration

$$\int f(\mathbf{z}; \boldsymbol{\mu}, \boldsymbol{\Sigma}) f(\mathbf{z}; \mathbf{m}, q\mathbf{S})^\alpha d\mathbf{z} = \frac{\exp\{-\alpha(\mathbf{m} - \boldsymbol{\mu})'(\mathbf{I} + \alpha\mathbf{W})^{-1}(\mathbf{m} - \boldsymbol{\mu})/2\}}{\det(\mathbf{I} + \alpha\mathbf{W})^{1/2} (2\pi)^{(\alpha p)/2} (\det q\mathbf{S})^{\alpha/2}},$$

where $\mathbf{W} = q\boldsymbol{\Sigma}\mathbf{S}^{-1}$. The integral with respect to the Dirac measure in the second term of (13) is $f(\mathbf{x}; \mathbf{m}, \mathbf{S})^{1-q}$. Therefore, for given $q < 1$, maximizing (13) is equivalent to maximize

$$\begin{aligned} & (1 - \epsilon) \frac{\exp\left\{-\frac{(1-q)}{2}(\mathbf{m} - \boldsymbol{\mu})'(\mathbf{I} + (1-q)\mathbf{W})^{-1}(\mathbf{m} - \boldsymbol{\mu})\right\}}{(2\pi)^{(1-q)p/2} \det(\mathbf{I} + (1-q)\mathbf{W})^{1/2} (\det q\mathbf{S})^{(1-q)/2}} \\ & + \epsilon \frac{\exp\left\{-\frac{(1-q)}{2}(\mathbf{x} - \mathbf{m})'\mathbf{S}^{-1}(\mathbf{x} - \mathbf{m})\right\}}{(2\pi)^{(1-q)p/2} \det(q\mathbf{S})^{(1-q)/2}}. \end{aligned} \quad (14)$$

Note that for $\|\mathbf{x}\| \rightarrow \infty$, (14) consists basically of the two ridges

$$\det(\mathbf{W})^{(1-q)/2} \epsilon \text{ at } \mathbf{m} = \mathbf{x}, \quad (15)$$

$$\frac{\det(\mathbf{W})^{(1-q)/2}(1-\epsilon)}{\det(\mathbf{I} + (1-q)q\mathbf{W})^{1/2}} \text{ at } \mathbf{m} = \boldsymbol{\mu}. \quad (16)$$

If the covariance is known, $\mathbf{S} = \boldsymbol{\Sigma}$ and breakdown occurs when the maximum is at $\mathbf{m} = \mathbf{x}$, i.e. $\epsilon > [1 + (1 + q - q^2)^{p/2}]^{-1}$. If $\mathbf{S} \neq \boldsymbol{\Sigma}$, the ridge at $\mathbf{m} = \mathbf{x}$ is larger at $\mathbf{m} = \boldsymbol{\mu}$ if and only if

$$\epsilon > [1 + \det(\mathbf{I} + q(1-q)\mathbf{W})^{1/2}]^{-1} \geq [1 + (1 + q(1-q)\lambda^*)^{p/2}]^{-1},$$

where $\lambda^* = \lambda_{\boldsymbol{\Sigma}}^{\max}/\lambda_{\mathbf{S}}^{\min}$, and $\lambda_{\boldsymbol{\Sigma}}^{\max}$, $\lambda_{\mathbf{S}}^{\min}$ are the maximum and minimum eigenvalues of $\boldsymbol{\Sigma}$ and \mathbf{S} , respectively. Therefore, if q is fixed and $\lambda_{\mathbf{S}}^{\min} \rightarrow 0$, the function ψ is unbounded for any $\epsilon > 0$, meaning that scale and location breakdowns occur for any amount of contamination. However, if we define a sequence of the tuning constants depending on p such that $q_p \rightarrow 0$, as p grows and $q_p(1-q_p)\lambda^* \approx c$ for some constant c , then breakdown occurs only when $\epsilon > [1 + (1 + c)^{p/2}]^{-1}$. Particularly, if $q_p = o(\lambda^*)$, $c \rightarrow 0$ and breakdown occurs when $\epsilon \geq 1/2$.

3.2. Computational aspects and choice of q

For a given $q \neq 1$, (5) can be viewed as a weighting process of the log-likelihood score. Consequently, a simple re-weighting algorithm is easily derived for computing the estimates. Let $s \in \{0, 1, \dots, s^*\}$ denote the iteration step.

1. If $s = 0$,
 $\boldsymbol{\mu}^{(s)}$ and $\boldsymbol{\Sigma}^{(s)}$ are set to be robust estimates of location and scale;

2. $\boxed{\text{For } 0 < s < s^*},$

$$\boldsymbol{\mu}^{(s)} = \sum_{i=1}^n v_i^{(s-1)} \mathbf{x}_i, \boldsymbol{\Sigma}^{(s)} = q^{-1} \sum_{i=1}^n v_i^{(s-1)} (\mathbf{x}_i - \boldsymbol{\mu}^{(s)})(\mathbf{x}_i - \boldsymbol{\mu}^{(s)})',$$

where

$$v_i^{(s)} = \frac{f(\mathbf{x}_i; \boldsymbol{\mu}^{(s)}, \boldsymbol{\Sigma}^{(s)})^{1-q}}{\sum_{i=1}^n f(\mathbf{x}_i; \boldsymbol{\mu}^{(s)}, \boldsymbol{\Sigma}^{(s)})^{1-q}}.$$

Particularly, if $q \neq 1$, the above procedure provides a relative-to-the-model downweighting. Observations that disagree sensibly with the model receive low weight. If $q = 1$, all the observations receive the same weight and the procedure is maximum likelihood estimation. For more details on general convergence behavior of re-weighting algorithms see Arslan (2004) and Maronna et al. (2006), p.331.

Different values of q correspond to estimators with different robustness and efficiency levels. Thus, having a reasonable strategy for selecting q is crucial. One approach is to choose $q < 1$ with largest empirical efficiency. We consider the ratio $\Lambda(q, \boldsymbol{\theta}_0, g) = \mathbf{V}_1(\boldsymbol{\theta}_0, g) \mathbf{V}_q^{-1}(\boldsymbol{\theta}_0, g)$, where \mathbf{V}_q is as in (6). Since $\boldsymbol{\theta}_0$ and g are unknown, in practice we consider a grid of distortion parameters $Q = \{q_1, \dots, q_r\}$ and compute the corresponding ML q E estimates $\hat{\boldsymbol{\theta}}_{q_1, n}, \dots, \hat{\boldsymbol{\theta}}_{q_r, n}$. Then, pick q^* such that $q^* = \max_{q \in Q} \text{tr}\{\hat{\Lambda}(q, \hat{\boldsymbol{\theta}}_{q, n}, dG_n)\}$ where G_n is the empirical distribution of the data. In a similar context, Windham (1995) pointed out the relationship between convergence rate of the estimates and empirical efficiency. Thus, he suggested a choice of the tuning parameter using an upper bound for empirical efficiency computed using the convergence rate of the estimates.

Typically, values of q between 1/2 and 1 work well: (i) For $q > 1$, the estimator has large bias. This is not surprising as the influence functions (11) and (12) are unbounded for such values. Conversely, for $q < 1$ the bias

decreases as q gets smaller. This is confirmed by our simulations. (ii) For q sufficiently small, usually smaller than $1/2$, the estimator gains considerable robustness. In this case, the estimator's variance increases sensibly. Moreover, the empirical convergence of the re-weighting algorithm above gets slower for values near or below $1/2$, especially when p gets larger. (iii) From our simulations in multivariate settings, as p gets large and $\det(\boldsymbol{\Sigma})$ gets smaller – as is the case for strongly correlated data – q near $1/2$ is required to maintain sufficient robustness.

4. Monte Carlo simulations

We perform an extensive simulation study in order to: (i) investigate the efficiency and robustness for various levels and types of contamination, dimensions of the parameter space and sample sizes (ii) evaluate the performance with respect to other well-known robust methods. Given a sample of size n , we generate B samples where about $(1 - \epsilon)n$ observations are from $N_p(\boldsymbol{\mu}_0, \boldsymbol{\Sigma}_0)$, while a smaller portion ϵn is from the contaminating distribution $N_p(\boldsymbol{\mu}_c, \boldsymbol{\Sigma}_c)$. To gauge performance, we compute the mean squared error with respect to $\boldsymbol{\mu}_0$ and $\boldsymbol{\Sigma}_0$.

We consider: (i) the ML q E when q is selected by the re-weighting method (see Section 3); (ii) the fully parametric MHDE computed using the ML q E with $q = 1/2$; (iii) the MLE; (iv) the semi-parametric MHDE based on nonparametric estimation (only for $p=1$ as the nonparametric analysis for choosing the bandwidth in higher dimensions is cumbersome and the convergence of $\hat{g}_h(\mathbf{x})$ has shown to be slower for $p > 1$); (v) the Huber Estimator, with re-descending

influence function (Campbell, 1980). Particularly, the MHDE uses non-parametric kernel density estimate $\hat{g}_h(\mathbf{x})$ to minimize the Hellinger Distance. We use Epanechnikov kernel with bandwidth $h = s_n c_n$, where $c_n = 0.5$ and $s_n = (0.6745)^{-1} \text{median}(|x_i - \text{median}(X_i)|)$ (e.g. see Bhandari et al. (2006)). Huber's estimator has been implemented as in Campbell (1980), so that the influence function is bounded and re-descending for large values of the Mahalanobis distance. This type of implementation allows direct comparisons with our estimator because the weights of extreme atypical observations tend to be zero (Hampel et al., 1986).

4.1. Univariate location and scale

We set $p = 1$, $\boldsymbol{\mu}_0 = 0$, $\boldsymbol{\Sigma}_0 = 1$ for $B = 1000$ Monte Carlo samples of size 100, 500 and 1000, with contaminations $\{5\%, 10\%, 20\%, 30\%\}$.

Contaminated location. Table 1 reports the mean squared errors for $\boldsymbol{\theta}$ when $\boldsymbol{\mu}_c = 2, 4, 6, 8$ and $\boldsymbol{\Sigma}_c = \boldsymbol{\Sigma}_0$. The MLqE with optimal q or $q = 1/2$ are always outperforming the MLE, MHDE and Huber estimators; when the contamination is small ($\epsilon = 5\%, 10\%$), the MLqE with $q = 1/2$ is performing best if the contaminating model is not located far from the true model ($\boldsymbol{\mu}_c = 2, 4$), while the MLqE with optimal q beats the other methods when the true and contaminated density are clearly distinct (e.g.: $\boldsymbol{\mu}_c = 6, \boldsymbol{\mu}_c = 8$). By increasing the level of contamination ($\epsilon = 20\%, 30\%$), the MLqE with $q = 1/2$ tends to outperform all other estimators since the larger percentage of outliers tends to increase the bias in the estimates and the distance between the contaminated and the true model plays a minor role.

Contaminated scale. Table 2 shows the mean squared errors for $\boldsymbol{\theta}$ when $\boldsymbol{\Sigma}_c = 2, 4, 6, 8$ and $\boldsymbol{\mu}_c = \boldsymbol{\mu}_0$. This type of contamination generates a strong

$\epsilon = 5\%$					
μ_c	MLE	MLqE	MLqE ($q = 1/2$)	MHDE	Huber
2	0.053 (0.001)	0.020 (0.000)	0.009 (0.000)	0.013 (0.000)	0.031 (0.001)
4	0.625 (0.005)	0.007 (0.001)	0.006 (0.000)	0.013 (0.000)	0.075 (0.002)
6	3.019 (0.014)	0.004 (0.000)	0.006 (0.000)	0.010 (0.000)	0.008 (0.000)
8	9.425 (0.033)	0.004 (0.000)	0.006 (0.000)	0.010 (0.000)	0.008 (0.000)
$\epsilon = 10\%$					
μ_c	MLE	MLqE	MLqE ($q = 1/2$)	MHDE	Huber
2	0.177 (0.002)	0.072 (0.001)	0.023 (0.001)	0.037 (0.001)	0.130 (0.002)
4	2.242 (0.011)	0.087 (0.006)	0.007 (0.000)	0.045 (0.001)	1.094 (0.010)
6	10.882 (0.034)	0.005 (0.000)	0.006 (0.000)	0.023 (0.001)	5.6353 (0.0208)
8	33.938 (0.082)	0.004 (0.000)	0.006 (0.000)	0.022 (0.000)	9.1128 (0.0379)
$\epsilon = 20\%$					
μ_c	MLE	MLqE	MLqE ($q = 1/2$)	MHDE	Huber
2	0.586 (0.004)	0.275 (0.002)	0.118 (0.001)	0.158 (0.001)	0.538 (0.004)
4	7.223 (0.025)	1.663 (0.008)	0.011 (0.000)	0.499 (0.003)	6.768 (0.025)
6	34.592 (0.078)	1.211 (0.066)	0.007 (0.000)	0.441 (0.010)	33.748 (0.078)
8	107.199 (0.189)	0.034 (0.017)	0.007 (0.000)	0.099 (0.001)	106.443 (0.191)
$\epsilon = 30\%$					
μ_c	MLE	MLqE	MLqE ($q = 1/2$)	MHDE	Huber
2	1.077 (0.006)	0.596 (0.003)	0.366 (0.003)	0.405 (0.002)	1.034 (0.006)
4	12.815 (0.039)	3.434 (0.007)	0.040 (0.003)	1.771 (0.006)	12.768 (0.039)
6	60.459 (0.124)	9.153 (0.019)	0.008 (0.000)	4.077 (0.012)	60.686 (0.125)
8	186.698 (0.285)	17.829 (0.053)	0.008 (0.000)	6.943 (0.023)	187.472 (0.286)

Table 1: Mean squared errors and standard errors (in parenthesis) for θ in univariate location contaminated models ($\mu_c = 2, 4, 6, 8$) and different levels of contamination ($\epsilon = 5\%, 10\%, 20\%, 30\%$) and $n=500$. Best results are in bold.

$\epsilon = 5\%$					
Σ_c	MLE	MLqE	MLqE ($q = 1/2$)	MHDE	Huber
2	0.031 (0.001)	0.006 (0.000)	0.006 (0.000)	0.008 (0.000)	0.010 (0.000)
4	0.632 (0.014)	0.005 (0.000)	0.006 (0.000)	0.009 (0.000)	0.012 (0.000)
6	3.332 (0.060)	0.005 (0.000)	0.006 (0.000)	0.010 (0.000)	0.011 (0.000)
8	10.756 (0.199)	0.005 (0.000)	0.006 (0.000)	0.010 (0.000)	0.011 (0.000)
$\epsilon = 10\%$					
Σ_c	MLE	MLqE	MLqE ($q = 1/2$)	MHDE	Huber
2	0.100 (0.002)	0.014 (0.000)	0.008 (0.000)	0.012 (0.000)	0.031 (0.001)
4	2.342 (0.032)	0.008 (0.000)	0.007 (0.000)	0.017 (0.000)	0.063 (0.001)
6	12.868 (0.163)	0.007 (0.000)	0.007 (0.000)	0.019 (0.000)	0.037 (0.001)
8	40.822 (0.528)	0.006 (0.000)	0.007 (0.000)	0.020 (0.000)	0.029 (0.001)
$\epsilon = 20\%$					
Σ_c	MLE	MLqE	MLqE ($q = 1/2$)	MHDE	Huber
2	0.377 (0.005)	0.048 (0.001)	0.016 (0.000)	0.030 (0.001)	0.161 (0.003)
4	9.203 (0.087)	0.016 (0.000)	0.013 (0.000)	0.059 (0.001)	0.576 (0.010)
6	49.431 (0.455)	0.012 (0.000)	0.010 (0.000)	0.069 (0.001)	0.505 (0.010)
8	163.059 (1.513)	0.010 (0.000)	0.008 (0.000)	0.071 (0.001)	0.312 (0.008)
$\epsilon = 30\%$					
Σ_c	MLE	MLqE	MLqE ($q = 1/2$)	MHDE	Huber
2	0.845 (0.009)	0.118 (0.002)	0.032 (0.001)	0.063 (0.001)	0.450 (0.006)
4	20.484 (0.162)	0.154 (0.013)	0.029 (0.001)	0.170 (0.002)	3.742 (0.049)
6	109.913 (0.826)	0.030 (0.006)	0.019 (0.001)	0.215 (0.002)	6.649 (0.141)
8	362.255 (2.662)	0.016 (0.000)	0.014 (0.000)	0.233 (0.002)	5.224 (0.190)

Table 2: Mean squared errors and standard errors (in parenthesis) for θ when considering univariate contaminated models ($\Sigma_c = 2, 4, 6, 8$) and different levels of contamination ($\epsilon = 5\%, 10\%, 20\%, 30\%$) and $n=500$. Best results are in bold.

overlap between the outliers and the main bulk of the data, which makes it hard to detect. The empirical results support the ML q E with optimally chosen q when ϵ is small, while the ML q E with $q = 1/2$ should be preferred when ϵ is large ($\epsilon = 20, 30\%$). This agrees with our theoretical findings related to the breakdown point. Inspecting more closely the results, we notice that when increasing the level of contamination ϵ , the automatic procedure for selecting q determines $q = 1/2$ as optimal value most of the time, but still not always. Although the MHDE can compete with the ML q E for small ϵ , the Huber estimator with re-descending influence function is always underperforming and explodes when ϵ is large.

4.2. Multivariate location/scale

We considered multivariate scale/location normal distributions with $p = 2, 5, 10, 20, 50$, $\epsilon = 5\%, 10\%, 20\%, 30\%$, $n = 100p$ and $B = 1000$. We set $\boldsymbol{\mu}_0 = \mathbf{0}$ and $\boldsymbol{\Sigma}_0$ is a $p \times p$ matrix with variances equal to 1 and covariances all equal to $\rho = 0.2$. Tables 3 and 4 show the results when $\boldsymbol{\mu}_c = -\mathbf{4}$ and $\boldsymbol{\Sigma}_c = 4\boldsymbol{\Sigma}_0$ and $\boldsymbol{\Sigma}_c = 8\boldsymbol{\Sigma}_0$, respectively. As in the univariate case, the ML q E with optimal q or $q = 1/2$ tends to outperform the other estimators for various p and ϵ . Huber is outperforming them only for a small level of contamination ($\epsilon = 5\%$) and with $p = 10, 20, 30$. When ϵ increases, the ML q E with $q = 1/2$ is clearly superior to all the others. However, we notice that the automatic procedure to choose the optimal q lead to identify $q = 1/2$ as optimal value quite easily when ϵ is small and the number of variates p increases. In fact, increasing p leads to datasets with observations that tend to gather around the mean. This makes the separation of outliers from the rest of the observations a difficult task. In such situations the ML q E has

remarkable performance also for a large number of variates and a high percentage of outliers. On the other hand, the task appears to be more difficult when ϵ and the number of variates is small and the outliers are overlapping with the clean data.

The simulations suggest that the ML q E can provide a valid alternative in robust estimation when compared with other estimators. In fact, the ML q E estimator with optimal selected q and with $q = 1/2$ can deal with a whole range of situations in terms of (i) degree of overlapping between the outliers and the data, (ii) fraction of outliers and (iii) number of variates. The multivariate analysis suggest that the ML q E with $q = 1/2$ should be preferred when the number of dimension p is large, the percentage of outliers ϵ is large and they are distant from the main bulk of data, as we expected also from the theoretical analysis on the breakdown point. On the other hand, when p and ϵ are small and outliers are only partially overlapping with the clean data, the use of the ML q E with iterative choice of the optimal q should be preferred.

5. Applications to financial data

Despite the counterfactual empirical evidence, mainly due to the presence of the so-called stylized facts, the gaussian model for asset returns is still appealing for its simplicity and a common choice among practitioners, as in the case of Markowitz's mean-variance portfolio allocation. Divergence-based methods, such as the ML q E, could then represent an interesting solution: (i) They allow to keep a relatively simple representation, as the normal one,

$\epsilon = 5\%$				
p	MLE	MLqE	MLqE ($q = 1/2$)	Huber
2	1.135 (0.015)	0.033 (0.001)	0.044 (0.001)	0.207 (0.003)
5	0.774 (0.004)	0.007 (0.000)	0.147 (0.001)	0.133 (0.001)
10	0.659 (0.002)	0.079 (0.000)	0.079 (0.000)	0.104 (0.000)
20	0.632 (0.001)	0.054 (0.000)	0.054 (0.000)	0.079 (0.000)
50	0.615 (0.000)	0.040 (0.000)	0.040 (0.000)	0.066 (0.000)
$\epsilon = 10\%$				
p	MLE	MLqE	MLqE ($q = 1/2$)	Huber
2	3.973 (0.037)	0.034 (0.001)	0.042 (0.001)	1.592 (0.018)
5	2.755 (0.010)	0.007 (0.000)	0.272 (0.002)	0.947 (0.005)
10	2.373 (0.005)	0.108 (0.001)	0.108 (0.001)	0.646 (0.002)
20	2.279 (0.002)	0.066 (0.000)	0.066 (0.000)	0.477 (0.001)
50	4.194 (0.003)	0.092 (0.001)	0.092 (0.001)	0.627 (0.001)
$\epsilon = 20\%$				
p	MLE	MLqE	MLqE ($q = 1/2$)	Huber
2	12.761 (0.084)	2.764 (0.148)	0.581 (0.016)	9.948 (0.071)
5	8.972 (0.025)	1.974 (0.033)	1.221 (0.006)	6.493 (0.020)
10	7.683 (0.010)	3.735 (0.006)	0.223 (0.001)	4.480 (0.008)
20	7.360 (0.005)	0.082 (0.000)	0.082 (0.000)	3.134 (0.004)
50	7.136 (0.005)	0.053 (0.001)	0.053 (0.001)	2.103 (0.002)
$\epsilon = 30\%$				
p	MLE	MLqE	MLqE ($q = 1/2$)	Huber
2	22.962 (0.122)	6.012 (0.297)	5.752 (0.083)	21.074 (0.117)
5	16.043 (0.035)	5.942 (0.099)	4.889 (0.019)	14.287 (0.033)
10	13.675 (0.016)	9.128 (0.012)	0.645 (0.003)	11.307 (0.015)
20	13.045 (0.008)	5.088 (0.004)	0.103 (0.000)	9.220 (0.006)
50	4.934 (0.429)	0.022 (0.002)	0.022 (0.002)	2.567 (0.223)

Table 3: Mean squared errors and standard errors (in parenthesis) for $\text{vech}(\boldsymbol{\theta})$ when considering multivariate contaminated models ($\boldsymbol{\Sigma}_c=4\boldsymbol{\Sigma}_0$, $\boldsymbol{\mu}_c = -4$, $\rho=0.2$) with different size p and different levels of contamination ($\epsilon = 5\%, 10\%, 20\%, 30\%$) and $n=10p$. Best results are reported in bold.

$\epsilon = 5\%$				
p	MLE	MLqE	MLqE ($q = 1/2$)	Huber
2	1.628 (0.026)	0.032 (0.001)	0.043 (0.001)	0.165 (0.003)
5	0.936 (0.006)	0.007 (0.000)	0.157 (0.001)	0.077 (0.001)
10	0.741 (0.003)	0.086 (0.000)	0.086 (0.000)	0.042 (0.000)
20	0.678 (0.001)	0.059 (0.000)	0.059 (0.000)	0.031 (0.000)
50	0.632 (0.001)	0.040 (0.000)	0.040 (0.000)	0.023 (0.000)
$\epsilon = 10\%$				
p	MLE	MLqE	MLqE ($q = 1/2$)	Huber
2	5.480 (0.060)	0.033 (0.001)	0.041 (0.001)	1.123 (0.015)
5	3.363 (0.015)	0.007 (0.000)	0.247 (0.001)	0.500 (0.003)
10	2.675 (0.007)	0.114 (0.000)	0.114 (0.000)	0.282 (0.001)
20	2.439 (0.003)	0.073 (0.000)	0.073 (0.000)	0.183 (0.000)
50	2.283 (0.002)	0.049 (0.000)	0.049 (0.000)	0.124 (0.000)
$\epsilon = 20\%$				
p	MLE	MLqE	MLqE ($q = 1/2$)	Huber
2	18.357 (0.144)	0.179 (0.049)	0.346 (0.008)	10.314 (0.095)
5	11.078 (0.039)	0.863 (0.053)	0.777 (0.004)	5.065 (0.022)
10	8.743 (0.016)	1.862 (0.045)	0.172 (0.001)	2.551 (0.006)
20	7.882 (0.008)	0.090 (0.000)	0.090 (0.000)	1.463 (0.002)
50	7.362 (0.008)	0.058 (0.000)	0.058 (0.000)	0.850 (0.001)
$\epsilon = 30\%$				
p	MLE	MLqE	MLqE ($q = 1/2$)	Huber
2	34.059 (0.218)	3.151 (0.277)	9.232 (0.253)	27.530 (0.193)
5	20.315 (0.055)	3.124 (0.151)	2.956 (0.012)	15.218 (0.045)
10	15.818 (0.023)	5.130 (0.052)	0.335 (0.001)	9.795 (0.016)
20	14.150 (0.012)	3.573 (0.004)	0.101 (0.000)	6.114 (0.006)
50	24.327 (0.026)	0.121 (0.001)	0.121 (0.001)	6.414 (0.008)

Table 4: Mean squared errors and standard errors (in parenthesis) for $\text{vech}(\boldsymbol{\theta})$ when considering multivariate contaminated models ($\boldsymbol{\Sigma}_c=8\boldsymbol{\Sigma}_0$, $\boldsymbol{\mu}_c = -4$, $\rho=0.2$) with different size p and different levels of contamination ($\epsilon = 5\%, 10\%, 20\%, 30\%$) and $n=10p$. Best results are reported in bold.

with a reliable fit to the data, instead of pursuing estimation of more complicated models, often leading to unstable estimates. (ii) Tuning the parameter q allows for a flexible treatment of time periods with different volatility regimes. When the volatility is low, the data are typically well approximated by the normal model, so choosing q near 1 gives efficient estimates. When the volatility is high, choosing q closer to $1/2$ prevents larger downward (or upward) movements of prices to inflate the bias of the estimates. Section 5.1 and 5.2 report the analysis of univariate and multivariate financial data, respectively. In particular, Section 5.2 shows how the ML q E could lead to build attractive investment strategies in a mean-variance framework.

5.1. Standard & Poor's 500 data

We apply our method to 1651 monthly observations of the log-return of the S&P 500 from January 1981 to August 2008. The returns are centered around zero, are leptokurtic with a longer left tail (the sample median, skewness and kurtosis are 0.0054, -0.3285 and 14.4584, respectively). In Table 5, we report the mean and standard deviation estimates computed by the different methods described in Section 4. The ML q E with optimally selected q and with $q = 1/2$ gives larger mean and smaller standard deviations estimates than those obtained using MLE and Huber. The semi-parametric MHD estimate of the mean value is the largest, while the standard deviation is between the ML q estimate with $q = 1/2$ and that with optimally chosen q .

In Figure 3, we show that the ML q E weights observations according to their closeness to the assumed model. Extreme observations (both positive and negative), which would overly affect the final estimates, are given nearly

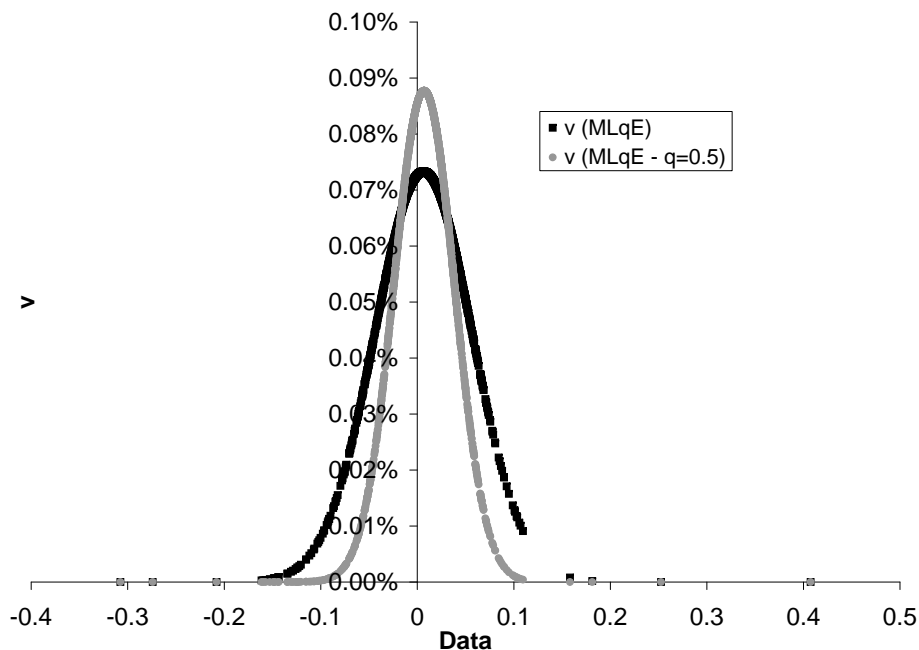


Figure 3: S&P500 Monthly Returns versus data weights.

zero weight. Since larger losses tend to occur more frequently – and with larger absolute value – than larger gains, the process of smooth re-weighting of all the observations results in larger expected returns and more conservative volatility estimates. The distribution of the data weights for $q = 1/2$ has skewness=-0.89 and kurtosis=-0.42, while the one for the weights for the optimally chosen q has skewness=-1.87 and kurtosis=3.21. This is not surprising given that setting $q = 1/2$, which is equivalent to minimize a version of the Hellinger distance, provides more conservative estimates by downweighting a larger amount of observations.

Finally, we compute a sequence of estimates for the expected returns

	ML q E	ML q E	MLE	Huber	MHDE
q	0.7	0.5	1	–	–
μ	0.0064	0.0071	0.0034	0.0043	0.0074
σ	0.0330	0.0314	0.0406	0.0342	0.0323

Table 5: Expected return (μ) and volatility (σ) estimates for monthly S&P 500 (January 1981 – August 2008).

and volatilities using moving windows of 120 observations and rolling ahead 12 months at the time. The estimates are computed by the ML q method with optimally chosen q . Figure 4 shows clearly that the automatic choice of the optimal q allows for a flexible treatment of periods characterized by high (low) volatility and by a large (small) proportion of anomalous data. Actually, q near 1/2 is automatically selected in time periods characterized by negative economic events, such as the end of the Postwar Boom (1920-1921), the Great Crash (1929-1932), World War II and the Postwar period, the first and second oil shocks (1973 and 1981), the Black Monday (1987) and the end of dot.com (2000-2002), etc. Therefore, a stronger robustification provides more reliable long-term estimates in stressful economic periods. This also successfully validates the procedure for automatic selection of q .

5.2. Mean-variance portfolio allocation

Markowitz (1952) opened the road to modern investment theory by introducing the mean-variance (MV) approach for optimally building portfolios of p assets. Assuming that asset returns are normally distributed with location vector $\boldsymbol{\mu}$ and scale $\boldsymbol{\Sigma}$, the portfolio expected return can be computed as $r_p(\mathbf{w}) = \mathbf{w}'\boldsymbol{\mu}$, while the portfolio expected variance is $\sigma_p^2 = \mathbf{w}'\boldsymbol{\Sigma}\mathbf{w}$, where \mathbf{w}

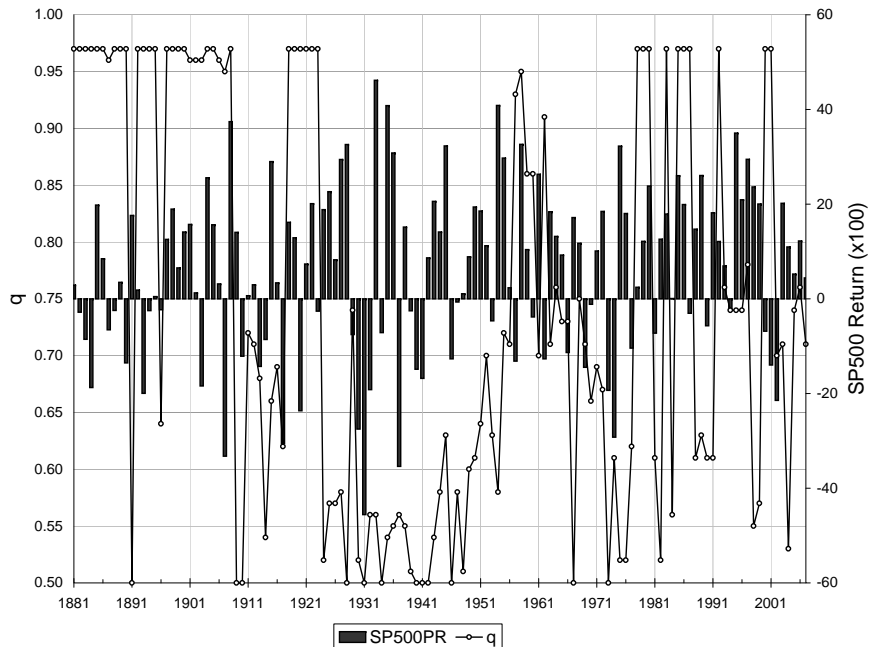


Figure 4: S&P500 Price Annual Returns and Optimal q Values (Period: December 1881-December 2007).

is the $p \times 1$ vector of asset weights. Markowitz's approach relies on determining the Pareto front of the optimal portfolios by solving a multiobjective optimization problem, where the targets are to minimize the portfolio's variance while maximizing the portfolio expected returns subject to the budget constraint $\mathbf{1}'\mathbf{w} = 1$, where $\mathbf{1}$ is a $p \times 1$ vector of ones. Then, for a given investor's risk aversion $\gamma > 0$, the optimization problem can be solved as

$$\arg \max_{\mathbf{w}} r_p(\mathbf{w}) - \frac{\gamma}{2} \sigma_p^2(\mathbf{w}) \quad \text{s.t.} \quad \mathbf{1}'\mathbf{w} = 1 \quad (17)$$

Different values of γ yield different investment strategies and determine

all the optimal portfolios which define the Pareto front or the so-called mean-variance efficient frontier. The optimal MV portfolio obtained by solving (17) is

$$\mathbf{w}_\gamma^*(\boldsymbol{\mu}, \boldsymbol{\Sigma}) = \gamma^{-1} \boldsymbol{\Sigma}^{-1} \left(\boldsymbol{\mu} - \frac{(\lambda_2(\boldsymbol{\Sigma}) - \gamma) \boldsymbol{\iota}}{\lambda_1(\boldsymbol{\Sigma})} \right). \quad (18)$$

where $\lambda_1 = \boldsymbol{\iota}' \boldsymbol{\Sigma}^{-1} \boldsymbol{\iota}$, $\lambda_2(\boldsymbol{\Sigma}) = \boldsymbol{\iota}' \boldsymbol{\Sigma}^{-1} \boldsymbol{\mu}$. The global minimum variance portfolio (GMV), which corresponds to the extreme risk aversion, is then $\mathbf{w}_\infty^* = \lim_{\gamma \rightarrow \infty} \mathbf{w}_\gamma^* = \boldsymbol{\Sigma}^{-1} \boldsymbol{\iota} / \boldsymbol{\iota}' \boldsymbol{\Sigma}^{-1} \boldsymbol{\iota}$.

The estimates of the optimal portfolio are defined by plugging-in the MLq estimates in Eq.(18) and computing $\widehat{\mathbf{w}}_q^* = \mathbf{w}_\gamma^*(\widehat{\boldsymbol{\mu}}_{q,n}, \widehat{\boldsymbol{\Sigma}}_{q,n})$. By continuity of $\widehat{\mathbf{w}}_q^*$, the properties for the MLqE of $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ discussed in Section 3 are readily extended to the plug-in estimator $\widehat{\mathbf{w}}_q^*$ as well (see details in Appendix).

5.2.1. Empirical Analysis

We analyze 339 monthly log-returns of 8 MSCI Indexes (USA, Japan, Pacific EX JP, France, Germany, Italy, Spain, UK) from January 1981 to April 2009. Estimates of the mean and covariance computed for different estimators are used to determine the Pareto front. In Figure 5, we show that the MLq estimates for the expected return are larger than ML estimates (left plot). Moreover, the MLqE gives more conservative estimates of the index variances (right plot). Huber robust estimates of the variance are similar to MLq estimates with $q = 1/2$, but closer to the ML estimates for the expected return.

In Figure 6, we show the Mahalobis distances of data using the MLq, ML and Huber estimates. The MLqE determines several points with relatively larger distances than for ML and Huber estimates. Thus, the MLqE smoothly

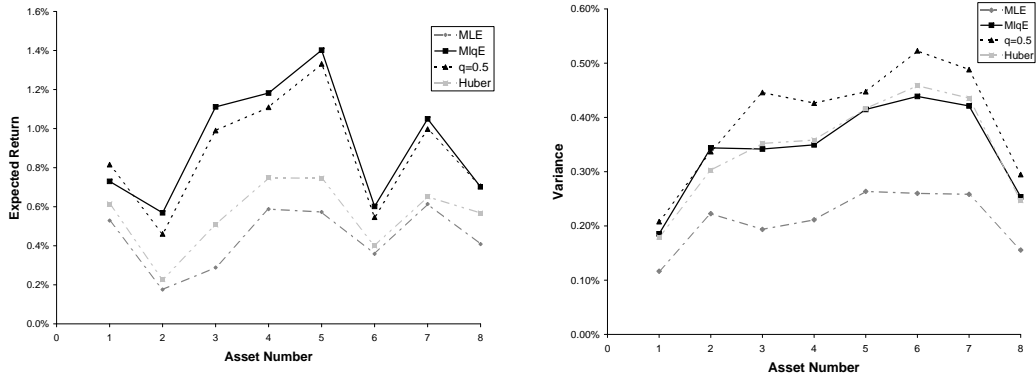


Figure 5: Expected Returns Estimates (left) and Variance Estimates (right) for the 8 MSCI Indexes (1: USA, 2: Japan, 3: Pacific ex Jp, 4: France, 5: Germany, 6: Italy, 7: Spain and 8: UK).

enhances the degree of outlyingness of each observation relative to the others, allowing for a finer detection of data inconsistent with the normal model. In Figure 7, we plot the efficient frontiers when short selling is allowed ($-1 \leq w_i \leq 1$, $i = 1, \dots, p$). The frontiers of the MLqE dominate those of ML and Huber, yielding portfolios with larger (in-sample) expected returns and smaller (in-sample) risk. Similar findings were obtained in case of no short selling.

We set-up a dynamic investment strategy and assess both in- and out-of-sample performances. We consider a rolling window scheme, where we hold the GMV portfolio and update its allocation every month using new estimates. The minimum variance portfolio is a typical choice as a benchmark for comparing different methods, since it is the least affected by the expected return estimates and possible large fluctuations due to the instability in the optimization process (Best and Grauer, 1991). The expected returns and

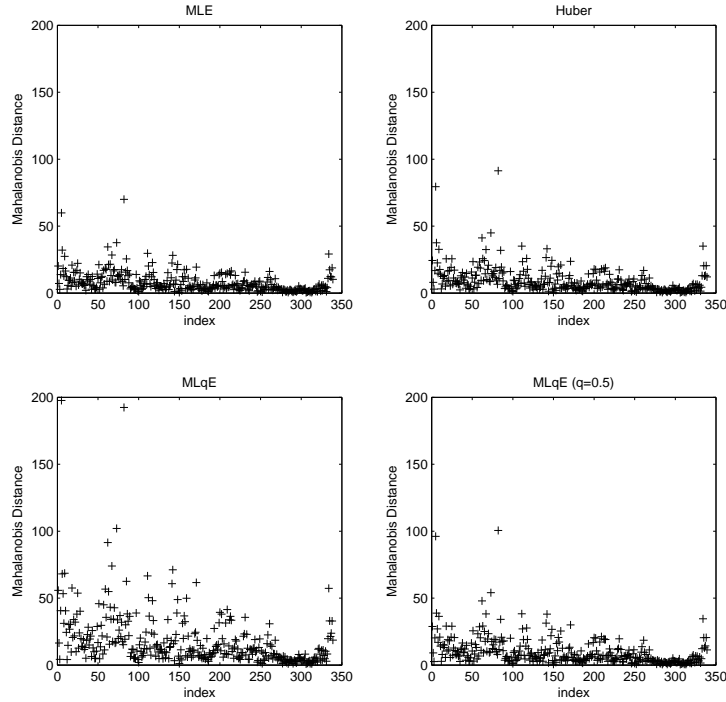


Figure 6: Mahalanobis Distances with respect to MLE, Huber, ML q E and ML q E ($q = 1/2$) estimates.

covariance matrix estimates were obtained by using the different methods on window of 60 observations. The out-of-sample performance is evaluated by computing the one-month-ahead portfolio return (for a total of 279 out-of-sample returns). The out-of-sample variance of the portfolio is computed using the optimal weights determined in-sample and covariance matrix estimate computed on 60 consecutive observations (including the out-of-sample 1-month-ahead observation). Figure 8 shows the boxplots of in- and out-of-sample returns and variances for GMV portfolios. The distributions of in-sample ML q returns are centered on larger values than those for MLE and

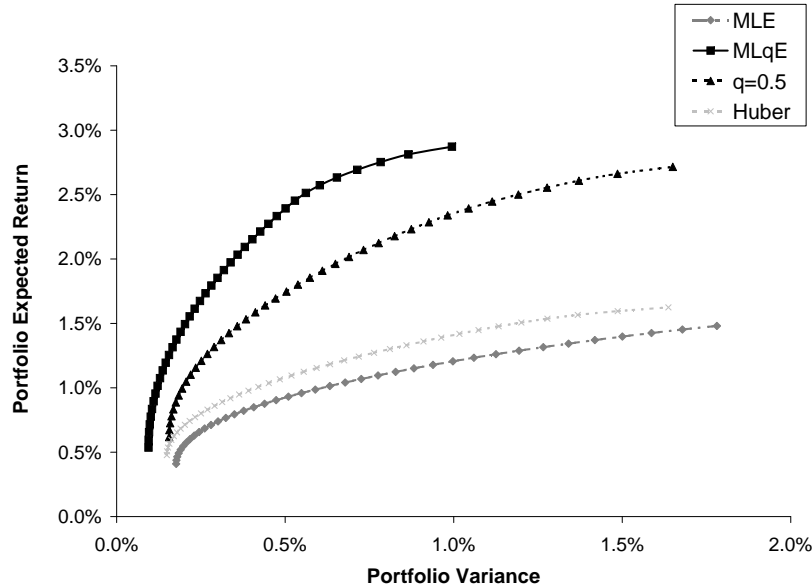


Figure 7: Mean-Variance Efficient Frontier with short selling. Period: January 1981- April 2009.

Huber, while the distributions of out-of sample returns are similar. Interestingly, however, the distributions of the portfolio variances for the $MLqE$ are much less spread for both in- and out-of-sample setups.

The $MLqE$ gives minimum-variance portfolios with slightly improved out-of-sample annualized mean returns, but also slightly larger volatility of the out-of-sample portfolio returns time series (see Table 6). The skewness and kurtosis coefficients of the out-of-sample returns, are much smaller for MLq estimates. Interestingly, this reveals that GMV portfolios are more robust to extreme fluctuations than those obtained by plugging-in the ML and Huber estimates. Hence, investor's trading strategies appear to be less affected by extreme risks.

	q	Annualized Volatility (%)	Annualized Mean Return (%)	Skewness	Kurtosis
ML q E	Opt.	1.094	0.170	-1.094	7.838
ML q E	1/2	1.137	0.171	-1.193	7.568
MLE	1	0.985	0.157	-1.521	9.345
Huber	-	1.104	0.154	-1.554	9.933

Table 6: Out-of-sample annualized volatility and mean return, skewness and kurtosis of the GMV portfolios using the ML q E, ML q with $q = 1/2$, MLE and Huber estimates.

6. Discussion and further research

In this work, we have studied parametric estimation based on minimization of the q -entropy and use it to estimate expected returns and volatilities of financial assets.

From a methodological viewpoint, the ML q E has several advantages for parametric density estimation: (i) Its behavior is characterized by well-established theoretical properties, which can be easily extended to the portfolio weights of the optimal allocation determined in a mean-variance framework. (ii) It provides a feasible way to use power-divergences and Hellinger distance, which otherwise would require nonparametric density estimation. All the complications of bandwidth selection and curse of dimensionality make the latter impracticable in many multivariate financial problems. (iii) The user can flexibly tune the trade-off between efficiency and robustness by a single parameter q . (iv) It can be easily implemented by a simple and fast procedure that automatically re-weights outliers depending on their closeness to the assumed model and also computes the optimal tuning parameter q .

Our theoretical findings and simulation results have shown that the $MLqE$ can handle different types and levels of contamination. The automatic choice of q treats conveniently a range of situations: q close to 1 should be preferred when the data are nearly clean in order to retain accuracy. These values also correspond to negligible loss of efficiency compared to the MLE. When a portion of observations disagree with the assumed model, moving q towards $1/2$ increases the robustness of the estimates.

The procedure works well when p is moderate. However, when the number of correlated variables is large compared to the sample size and q is far from 1, the algorithm in Section 3.2 may give sub-optimal results. In our view, more work to make computations feasible when p is large would be valuable. Moreover, inspecting the performance of other methods for optimal selection of tuning parameter q is of order. For example, one could use measures of the worst-case scenario bias (gross error sensitivity) or minimize approximations of the mean squared error under an ϵ -amount of contamination. Other strategies such as computing bootstrap estimates the mean squared error of the estimates or standard cross-validation methods should be explored as well.

Our empirical investigation on financial data has shown that robust divergence-based methods, such as the $MLqE$, are appealing for providing a reliable fit to real world data while keeping simple models, as the normal one. This could be a valuable alternative to unstable estimates of complex models. Our analyses show that the $MLqE$ works well in presence leptokurtic data, asymmetry gain/losses and different volatility regimes.

The analysis of S&P500 and of the classic mean-variance portfolio allo-

cation strategy, have pointed out that our long-term investment strategies are less exposed to extreme risks and can be used to detect switches among volatility regimes. In fact, tuning the parameter q allows for a flexible treatment of different time periods: when the volatility is low, the data are typically well approximated by the normal model, so choosing q near 1 gives efficient estimates; when the volatility is high, choosing q closer to $1/2$ prevents larger downward (or upward) movements of prices to inflate the bias of the estimates. Further applications on different data, time-horizons and on a larger number of assets with more realistic trading strategies (e.g. including transaction costs) are of high priority in our agenda. Finally, although the focus here is on unconditional multivariate normal models, ML q estimation can be readily extended to richer stochastic representations for modeling time-dependency of the observations.

Appendix A. Properties for the MV portfolio

Appendix A.1. Asymptotics

By continuity of $\widehat{\mathbf{w}}_q^*$, the properties for the ML q E of $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ discussed in Section 3 can be extended to the plug-in estimator $\widehat{\mathbf{w}}_q^*$ as well. Given observations $\mathbf{x}_1, \dots, \mathbf{x}_n$ from g , applying the multivariate Delta Method (e.g. see Van der Vaart (1998)), we have that any consistent sequences of estimates $\widehat{\boldsymbol{\mu}}_{q,n}$ and $\widehat{\boldsymbol{\Sigma}}_{q,n}$, yield

$$\sqrt{n}(\widehat{\mathbf{w}}_{q,n}^* - \mathbf{w}^*) \xrightarrow{\mathcal{D}} N_p(\mathbf{0}, \mathbf{D}(\boldsymbol{\theta}_0, G)), \quad \text{as } n \rightarrow \infty, \quad (\text{A.1})$$

with asymptotic variance

$$\mathbf{D}(\boldsymbol{\theta}, G) = [\nabla_{\boldsymbol{\theta}} \mathbf{w}_{\gamma}^*(\boldsymbol{\theta})]' V_q(\boldsymbol{\theta}, G) [\nabla_{\boldsymbol{\theta}} \mathbf{w}_{\gamma}^*(\boldsymbol{\theta})] \quad (\text{A.2})$$

where \mathbf{V}_q can be computed by using the unbiased estimate (3.1.1). Moreover, $\nabla_{\theta} \mathbf{w}_{\gamma}^*(\boldsymbol{\theta})$ is a $p \times k$ block diagonal matrix with the following diagonal blocks. Each block can be computed using

$$\nabla_{\boldsymbol{\mu}} \mathbf{w}^* = \gamma^{-1}(\mathbf{I} - \mathbf{w}_{\infty}^* \boldsymbol{\iota}') \boldsymbol{\Sigma}^{-1}. \quad (\text{A.3})$$

and the derivative of \mathbf{w}^* with respect to $\text{vech}(\boldsymbol{\Sigma})$ is

$$\begin{aligned} \nabla_{\text{vec}(\boldsymbol{\Sigma})} \mathbf{w}^* &= \{\gamma^{-1}[(\boldsymbol{\mu}' \otimes \mathbf{I}) - \mathbf{w}_{\infty}^* (\boldsymbol{\mu}' \otimes \boldsymbol{\iota}')] \\ &\quad + \lambda_1^{-1}(\gamma^{-1} \lambda_2 - 1)[(\boldsymbol{\mu}' \otimes \mathbf{I}) - \mathbf{w}_{\infty}^* (\boldsymbol{\iota}' \otimes \boldsymbol{\iota}')] \} (\boldsymbol{\Sigma}^{-1} \otimes \boldsymbol{\Sigma}^{-1}). \end{aligned} \quad (\text{A.4})$$

Appendix A.2. Bounded influence function

As far as concerns robustness, boundness of the influence function for scale and location and global robustness can be easily transferred to the plug-in estimator. Recall that computing an IF for $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ entails differentiating the functional $T(g_{\epsilon})$ with respect to the contamination ϵ and evaluate at $\epsilon = 0$. Thus, to compute the IF for the MV portfolio, chain-differentiating $\mathbf{w}_{\lambda}^*(T(G_{\epsilon}))$ with respect to the contamination ϵ and evaluating at zero gives

$$IF(\mathbf{x}, \widehat{\mathbf{w}}_{q,n}^*) = -\boldsymbol{\Sigma}^{-1} IF(\mathbf{x}; \widehat{\boldsymbol{\Sigma}}_{q,n}) \mathbf{w}_{\gamma}^* + \gamma^{-1} \boldsymbol{\Sigma}^{-1} \{IF(\mathbf{x}; \widehat{\boldsymbol{\mu}}_{q,n}) \quad (\text{A.5})$$

$$+ \lambda_1^{-1} [\boldsymbol{\iota}' \boldsymbol{\Sigma}^{-1} IF(\mathbf{x}; \widehat{\boldsymbol{\Sigma}}_{q,n}) \boldsymbol{\Sigma}^{-1} \boldsymbol{\mu} - \boldsymbol{\iota}' \boldsymbol{\Sigma}^{-1} IF(\mathbf{x}; \widehat{\boldsymbol{\mu}}_{q,n})] \boldsymbol{\iota} \quad (\text{A.6})$$

$$- \lambda_1^{-2} (\lambda_2 - \gamma) [\boldsymbol{\iota}' \boldsymbol{\Sigma}^{-1} IF(\mathbf{x}; \widehat{\boldsymbol{\Sigma}}_{q,n}) \boldsymbol{\Sigma}^{-1} \boldsymbol{\iota}] \boldsymbol{\iota}. \quad (\text{A.7})$$

since \mathbf{x} appears only in the IF terms of the above expression. For $q < 1$, since both IFs for scale and location are bounded, the IF for the investment strategy is bounded as well. However, this is not the case if $q = 1$, because the asymptotic bias can be infinitely large when serious deviations from model assumptions occur.

- Arslan, O., 2004. Convergence behavior of an iterative reweighting algorithm to compute multivariate m-estimates for location and scatter. *Journal of Statistical Planning and Inference* 118 (1-2), 115 – 128.
- Basu, A., Lindsay, B. G., 1994. Minimum disparity estimation for continuous models: Efficiency, distributions and robustness. *Annals of the Institute of Statistical Mathematics* 46, 683–705.
- Bauwens, L., Laurent, S., Rombouts, J. V. K., 2006. Multivariate garch models: A survey. *Journal of Applied Econometrics* 21, 79–109.
- Beran, R., 1977. Minimum Hellinger distance estimates for parametric models. *The Annals of Statistics* 5, 445–463.
- Best, M. J., Grauer, R. R., 1991. On the sensitivity of mean-variance efficient portfolios to changes in asset means: some analytical and computational results. *Review of Financial Studies* 4 (2).
- Bhandari, S. K., Basu, A., Sarkar, S., 2006. Robust inference in parametric models using the family of generalized negative exponential disparities. *Australian & New Zealand Journal of Statistics* 48 (1), 95–114.
- Campbell, N., 1980. Robust procedures in multivariate analysis i: Robust covariance estimation. *Applied Statistics* 29, 231–237.
- Cont, R., 2001. Empirical properties of asset returns: stylized facts and statistical issues. *Quantitative Finance* 1, 223–236.
- DeMiguel, V., Nogales, F. J., 2009. Portfolio selection with robust estimation. *Operations Research* 57 (3), 560–577.

- Ferrari, D., La Vecchia, D., 2009. A fully parametric approach to minimum power-divergence estimation. Submitted to *Biometrika*.
- Ferrari, D., Yang, Y., 2009. Maximum L_q-likelihood estimation. To appear in *Annals of Statistics*.
- Hampel, F. R., Ronchetti, E., Rousseeuw, P., Stahel, W., 1986. *Robust Statistics: the Approach Based on Influence Functions*. Wiley, New York.
- Havrda, J., Charvát, F., 1967. Quantification method of classification processes: Concept of structural entropy. *Kibernetika* 3, 30–35.
- Huber, P. J., 1981. *Robust Statistics*. John Wiley & Sons.
- Lauprete, G., Samarov, A., Welsch, R., 2002. Robust portfolio optimization. *Metrika* 25, 139–149.
- Lindsay, B. G., 1994. Efficiency versus robustness: The case for minimum Hellinger distance and related methods. *The Annals of Statistics* 22, 1081–1114.
- Markowitz, H. M., 1952. Mean-variance analysis in portfolio choice and capital markets. *Journal of Finance* 7, 77–91.
- Maronna, R. A., Martin, R. D., Yohai, V. J., 2006. *Robust statistics : theory and methods*. John Wiley & Sons.
- Perret-Gentil, C., Victoria-Feser, M.-P., 2005. Robust mean-variance portfolio selection. Available at SSRN: <http://ssrn.com/abstract=721509>.

- Tsallis, C., 1988. Possible generalization of Boltzmann-Gibbs statistics. *Journal of Statistical Physics* 52 (1-2), 479–487.
- Van der Vaart, A. W., 1998. *Asymptotic Statistics*. Cambridge University Press, New York.
- Vaz-de Melo, B., Camara, R. P., 2005. Robust multivariate modeling in finance. *International Journal Of Managerial Finance* 4 (1), 12–23.
- Welsch, R. E., Zhou, X., 2007. Application of robust statistics to asset allocation models. *RevStat* 5 (1), 97–114.
- Windham, M. P., 1995. Robustifying model fitting. *Journal of the Royal Statistical Society, Series B: Methodological* 57, 599–609.
- Zhao, Z., 2008. Parametric and nonparametric models and methods in financial econometrics. *Statistical Surveys* 2, 1–42.

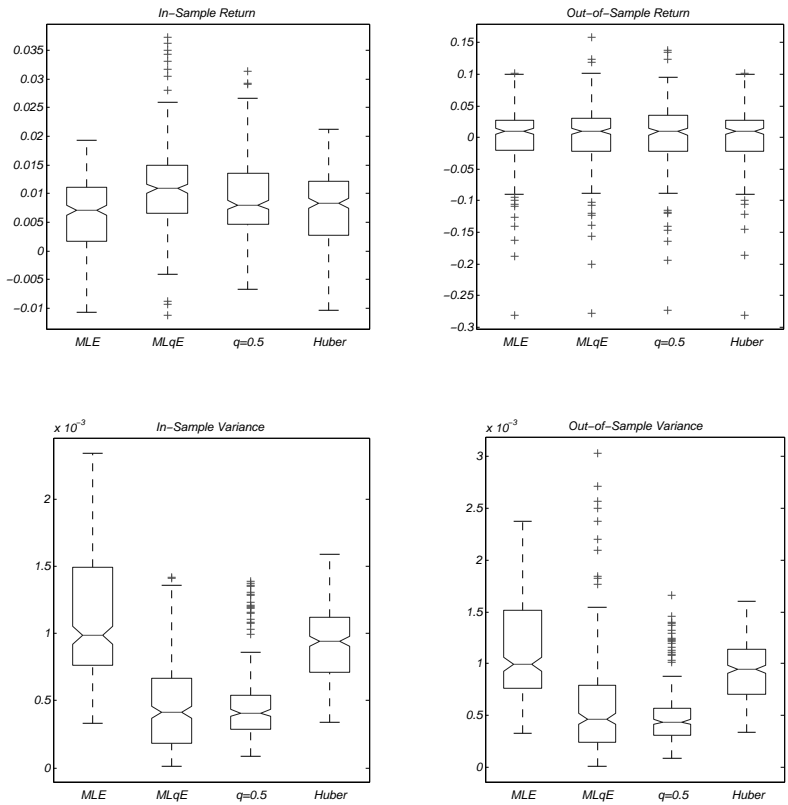


Figure 8: Boxplots of in-sample and out-of-sample GMV Portfolio Returns and Variance.